Standard Update

Better Data Through Standards

EPA

Spring 1999

Welcome to the first issue of ***Standard Update****, EPA's newsletter about data standards. In each issue, we will cover data standards-related topics—information you can use whatever your interest in information management may be. Feel free to read and reuse the information to suit your needs. For more information on any newsletter topic, see the Environmental Data Registry (EDR) web site at http://www.epa.gov/edr/. The EDR web site will also be your source for* ***Standard Update*** *back issues and a mechanism for reader comments.*

What's a Standard? You'll Know It When You *Don't* See It

Have you ever reached for a container of milk at the store and noticed it looks less full than the others? This probably doesn't happen very often, thanks, in part, to standards for weights and measurements. When standards are in place, we don't notice them. We buy the milk confident that it is the amount indicated on the label.

So, what is a standard? A standard can be defined as something agreed upon as a model, example, or rule for the measure of quantity, weight, extent, value, or quality. Standards affect almost every facet of our lives. Without standards, processes would become unreliable, commerce would stall, products would not work together, and consumer interests would go unprotected.

Historically, standards have been used for many years. The use of standards was documented in the Civil War, when ammunition sizes had to be standardized so that soldiers on the battlefield could share ammunition between muskets. And, at one time, the diversity of screw threads for the same application caused maintenance problems and posed a significant barrier to international trade. A global solution was supplied in the International Organization for Standardization (ISO) standards for metric screw threads.

Better Data, Greater Access

But, what do milk cartons and screws have to do with data standards? Standards–whether they are for screws or data–increase the reliability and effectiveness of many things we use. Standardized data is essentially more meaningful, more powerful, and easier to store. The benefits of data standardization are known and documented:

- Improved data quality.
- Increased data compatibility.
- Improved consistency and efficiency of data collection.
- Reduced data redundancy.
- Improved data access.

The key elements of a data standard are data element names, definitions, data types, and formatting rules.

Data standards are documented agreements on representations, formats, and definitions of common data.

Generally, data standards include text that describes procedures, implementation guidelines, and usage requirements. Additionally, standards may enable electronic reporting, data transfer protocols, or other information that facilitates and promotes widespread use. Regardless of the standard's composition, a data standard should be useable.

The true test of a data standard's effectiveness is its widespread acceptance and use.

EPA and Data Standards

EPA recognizes the benefit of data standardization. The Agency is developing methods to improve sharing environmental data across program systems, reporting regulatory and compliance information, providing user access, and integrating and storing environmental data. In support of Agency-wide data standards efforts, the Enterprise Information Management Division (EIMD) created the Environmental Data Registry (EDR). (http://www.epa.gov/edr/) In the EDR you will find data elements for representing items such as mailing and physical addresses, latitude and longitude, country codes, and calendar dates. The EDR is the central source of metadata (data that describes and defines other data) about Agency data and a source of well-formed data elements and value domains. The EDR is also a vehicle for reusing data elements from other data standards-setting organizations.

Standards Status

As part of the Reinventing Environmental Information (REI) initiative, EPA has committed to implementing an initial set of six data standards listed below. The standards include groups of data elements that relate to each of the subject areas and associated business rules for using data elements. The current status of each REI standard is presented below. For more information on these standards, see the EDR web site.

Date–Final. EPA's standard for calendar date is an 8-digit sequence composed of numeric characters in the format YYYYMMDD.

Economic Classifications (SIC/NAICS)–Final. This group includes ways to classify business activities, including industry classifications, product classifications, and product codes.

Latitude/Longitude–Interim. The group of data elements used for recording horizontal and vertical coordinates and associated metadata that define a point on earth.

Facility Identification–Interim. Provides a set of core data elements that properly identifies the location, the basic type/function of a regulated entity, and the association among multiple records relating to that entity.

Biological Taxonomy–Interim. Provides a common and consistent way to represent biological organisms in the collection, analysis, and exchange of environmental data.

Chemical Identification–Interim. Provides a common and consistent way to represent chemical substances regulated or monitored by EPA.

Each of these standards reflects Agency information needs as well as requirements for compliance with governmental, national, and international standards.

For additional information on Agency data standards activities, contact Marian Cody via e-mail at cody.marian@epa.gov or by phone at (202)260-5110.